1.Get and load a dataset. Study the structure of the dataset .Visualize all statistical measures like- mean, median, mode, range, Inter quartile range, Histograms, boxplots and scatterplots.

▾ Importing Required packages

```
import seaborn as sns
import pandas as pd
import numpy as np
import  matplotlib.pyplot  as  plt


from google.colab import drive
drive.mount('/content/drive')


#Loading the 'Titanic' data set
titanic   = pd.read_csv("/content/Titanic Data Set.csv")


#Printing the first 5 rows from the data set
titanic.head()
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN

```
#Printing the basic info about the data set
titanic.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 12 columns):
 #    Column       Non-Null Count   Dtype
---   ------       --------------   -----
 0    PassengerId  891 non-null     int64
 1    Survived     891 non-null     int64
 2    Pclass       891 non-null     int64
 3    Name         891 non-null     object
 4    Sex          891 non-null     object
 5    Age          714 non-null     float64
 6    SibSp        891 non-null     int64
 7    Parch        891 non-null     int64
 8    Ticket       891 non-null     object
 9    Fare         891 non-null     float64
 10   Cabin        204 non-null     object
 11   Embarked     889 non-null     object
dtypes: float64(2), int64(5), object(5)
memory usage: 83.7+ KB
```
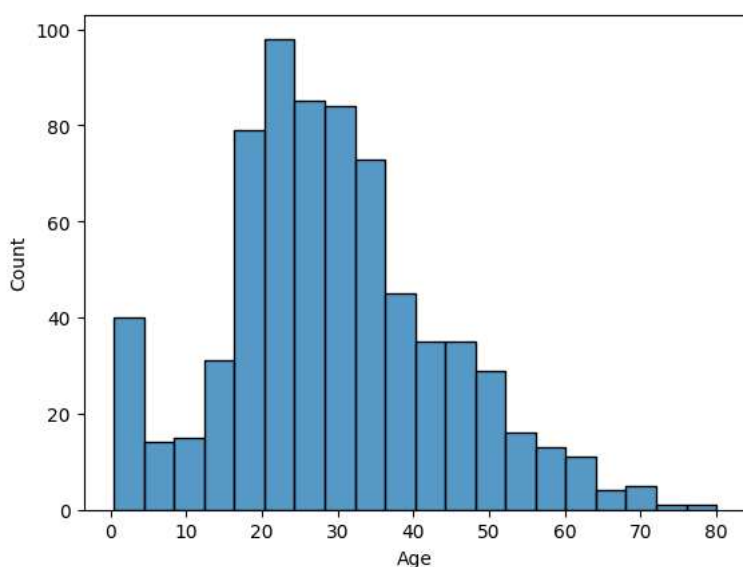
```
#Printing the basic info about the data set
titanic.describe()
```

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare

```
#Checking for null values
titanic.isnull().sum()
```

```
PassengerId      0
Survived         0
Pclass           0
Name             0
Sex              0
Age            177
SibSp            0
Parch            0
Ticket           0
Fare             0
Cabin          687
Embarked         2
dtype: int64
```

```
#Plots to find relationships between the variables
sns.histplot(titanic['Age'])
plt.show()
```


\# The above Histogram represents the count of people of different age groups.

From this graph,I can observe that most of the passengers belongs to 20-30 age group and a very less no. of passengers belongs to 60-80 age group.

```
sns.displot(x ="Age",kde=True,bins = 5 ,hue = titanic["Survived"] , palette = 'Set1', data=titanic)
plt.show()
```

200 —
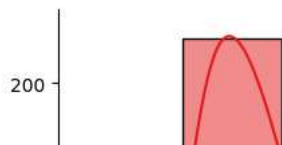
The above graph represents the relation between the age groups of the people and their Survival state

From this graph, I can observe that both most no. of the passengers Survived and most no. of passengers Dead belongs to the age group of 20-30, as most of the passengers are of that age group..!

```
sns.catplot(x ="Sex", hue ="Survived",kind ="count", data = titanic)
plt.show()
```
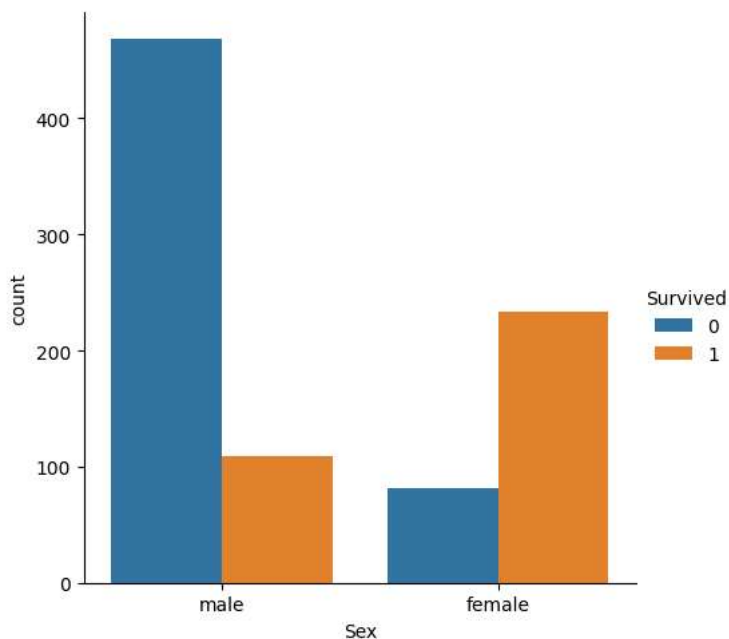

The above bar graph represent the Survival rate of people based on their Gender. From this plot, I can say that most of the males died and most of the females survived.

```
sns.violinplot(x ="Sex", y ="Age", hue ="Survived",data = titanic, split = True)
plt.show()
```
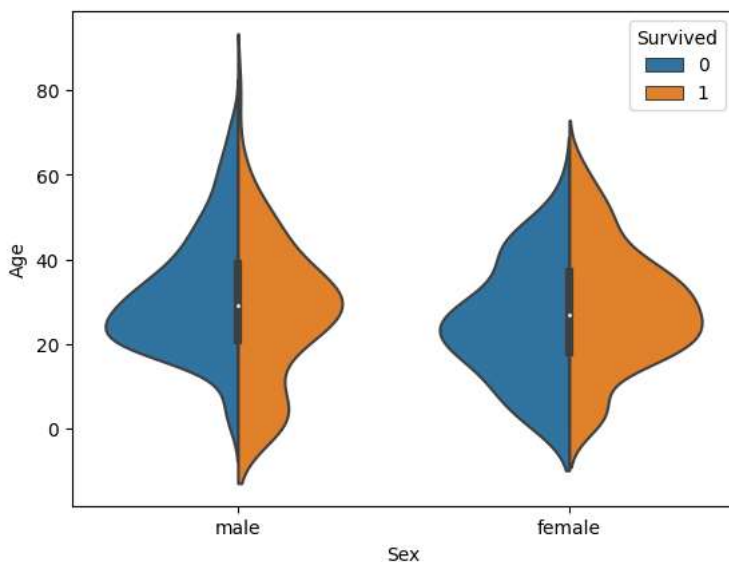

The above plot represent the relation between Survival rate and age groups of the passengers.

From the above plots, I can Observe that both Survival rate and death rate is more in the age group of 20-30 as most of the passengers belongs to that group.

```
sns.catplot(x ="Embarked", hue ="Survived", col ="Sex", kind="count",data = titanic)
plt.show()
```
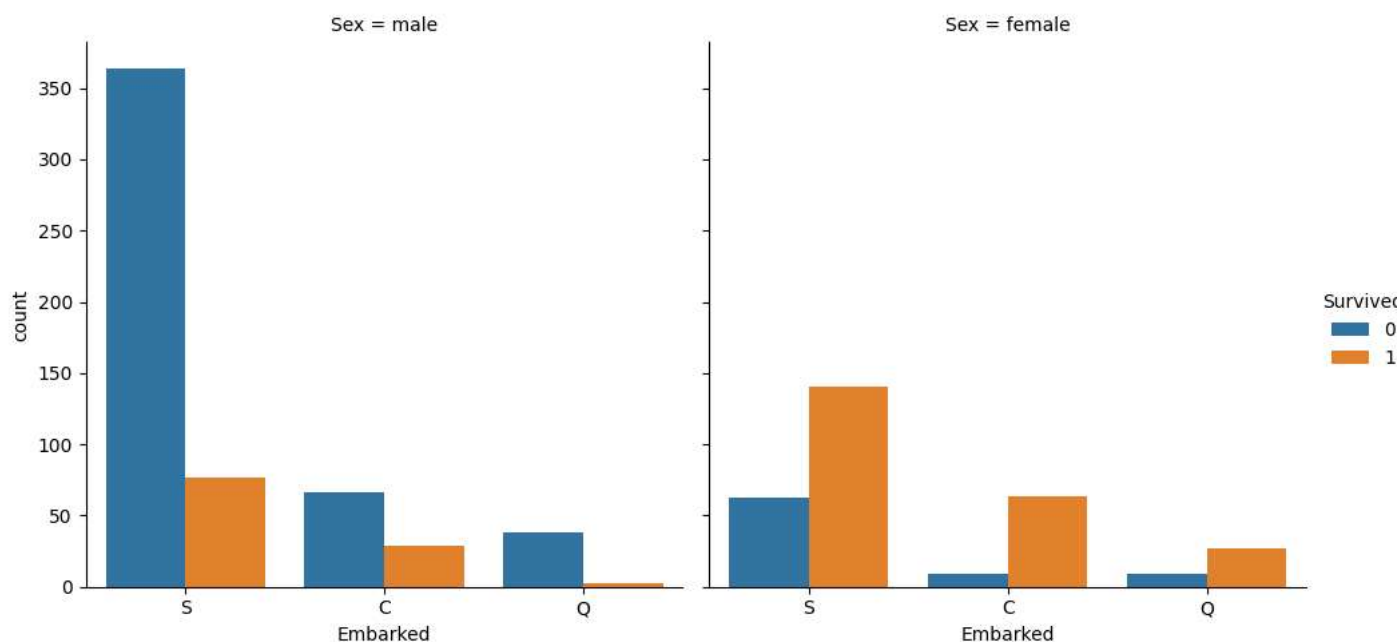

The above sub plots represent the relation between count of passengers embarking at each station and Survival status according to thir Gender.

From the above plot, I can observe that more no. of passengers were embarking at station S and less no. of passengers at station Q of both genders. So, The survival rate and death rate is also high at Station S. Also , more than half of the male passengers died and more than half of female passengers Survived at all stations.

```
one=two=three=0
for i in titanic["Pclass"]:
    if(i==1):
        one+=1
    elif(i==2):
        two+=1
    elif(i==3):
        three+=1
data = [one,two,three]
keys = ['Class 1', 'Class 2', 'CLass 3']
palette_color = sns.color_palette('bright')
plt.pie(data, labels=keys, colors=palette_color, autopct='%.0f%%')
plt.show()
```

\# This is a pie chart which represents the percentages of passengers who travel in different classes. From this, I can say that many passengers travelled in class 3

```
sns.countplot(x='Pclass', hue = 'Survived', data = titanic,palette = 'Set1')
plt.show()
```
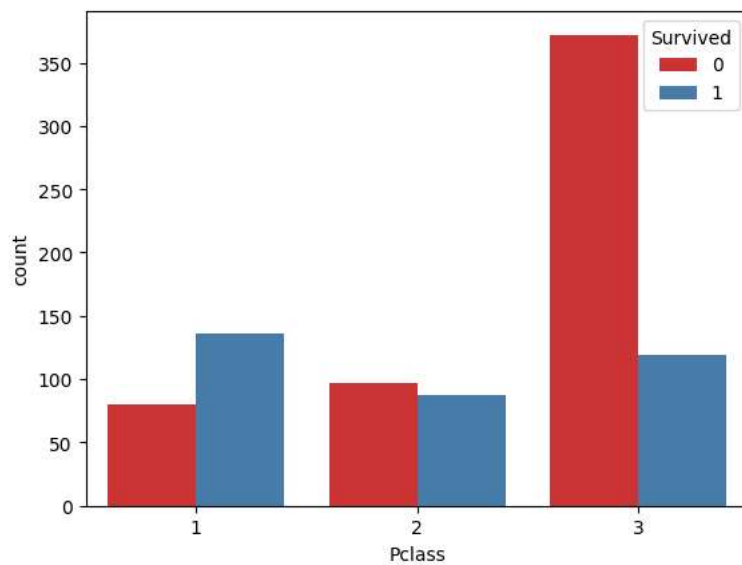

\# The above represents the relation between Survival rate and Pclass.

From this, I can say that more no. of passengers died who travelled in class 3.

Conclusion:

\# From the above Graphs , I can say that whether a passenger Survived or not depends upon the Features Pclass, Age, Cabin, Embarked.